Bachelor Thesis

Analysis of Daily Activity Patterns in Psychiatric Patients

Bachelor in Communication System Engineering

Author:
Belén Vázquez Pastor

Supervisor:
Alfredo Nazábal Rentería

Tribunal

Presidente: Carmen Peláez Moreno

Vocal: José Francisco Algorri Genaro

Secretario: José Carlos Pulido Pascual

Realizado el acto de defensa y lectura del Trabajo de Fin de Grado el día 6 de Julio de 2016 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO PRESIDENTE

Acknowledgements

I would like to express my gratitude to my thesis supervisor, Alfredo Nazábal Rentería, for his support and suggestions.

To my parents, for their trust and patience, for their amazing moral support so needed at this difficult stage just weeks away of finishing my degree.

To my classmates. Thank you for staying together through all these classes and years, thank you for your understanding and support.

And finally, I want to dedicate this thesis to my late friend Viviana. I hope this work may help others in the same way as it could have helped you.

Thank you all.

Disclaimer

Figure [3] and the original input data of the daily activity patterns collected from the patients have been handed down to me by my supervisor, they do not belong to me. For confidentiality reasons, patient data has been collected, handed and processed anonymously to avoid connection to any real person.

Figures [4a][4b] belong to Shimmer®, property of Realtime Technologies Ltd.[1]

Abstract

The recognition of human activities or *Human Activity Recognition* (most commonly abbreviated as *HAR*) is a research field of great interest for security, medical and military applications, even positioning services. Some medical application examples include patients in need of monitoring such as the elderly and others in need of physical therapy as part of their treatment [2]. In particular, it is a very appealing proposal to monitor the behavior of patients with dementia or other brain disorders in order to identify abnormal activities and avert unwanted consequences [3].

This thesis proceeds with the representation of the daily activity patterns of psychiatric patients with the goal of storing, selecting and analyzing patient data as a tool to aid doctors. To this aim, all data gathered in prior stages of the project will be processed according to the parameters introduced by the end user. Furthermore, this work will provide a graphical user interface where doctors can obtain different representations of the activity patterns of their patients to aid them evaluate their behavior and improve their treatment.

Contents

List of Figures

1 Introduction

1.1 Purpose

This bachelor thesis proposes the representation of daily activity patterns obtained from psychiatric patients. To accomplish this feature, a graphical user interface must be implemented to visualize all this data.

To ease the interpretation of this raw data, data clarity is a must: Standard graphs and standard formats are required to avoid distractions from the important patterns that may appear in the representation.

1.2 Motivation

Psychiatric patients suffering from disorders like depression or schizophrenia develop activity patterns that show their psychological state. The monitoring of those activity patterns is a valuable tool that doctors can employ to evaluate the health conditions of their patients and decide their treatment. This monitoring could be achieved through different types of devices that should be unobtrusive and easy to wear for the patients and easy to handle by the medical staff.

Today, the rise of wearable devices such as *smartwatches* (Pebble Time, Moto 360, Apple Watch...) and other attachable gadgets (fitbit, Samsung Gear Fit...) provides an excellent opportunity to introduce sensors in medical applications. Patients can be monitored using *wearable* sensors to proportionate accurate daily activities data whether they are in a restricted medical environment or in ambulatory conditions. In people like psychiatric patients suffering from bipolar disorders, depression or schizophrenia offers the possibility of carrying out rehabilitation from home and can be remotely monitored without the need of being hospitalized.

Moreover, public acceptance of other types of wearable devices such as heart rate monitors and fitness trackers indulges this train of thought. As they are no longer only associated to medical purposes, like Holter monitors or insulin pumps, anyone is able to wear a sensor without psychological prejudices or physical disturbances.

1.3 Contributions

This thesis main contribution is introduces the processing of the patient data provided by prior stages of the project. The data provided are the posterior probabilities of the activities performed by the patients given the observations from the wearable sensors, at a sample rate of 16Hz. This data will be analyzed in different conditions. We will consider slots of one minute of activity when we study the patterns in short time intervals of one hour. Also, we will use slots of 15 minutes when we represent full day activity patterns.

The statistical analysis will provide a more meaningful insight into the data significance. Concepts as average behavior or mode behavior will be displayed for medical evaluation. The graphical representation of the results will be provided in a standardized daily format (always from 8:00-24:00 hours), where different patient activities will be displayed together for comparison analysis.

A graphical interface is created with the purpose of simplifying the procedure of selecting and visualizing the daily activity patterns of psychiatric patients. The interface will allow doctors to select a specific patient, a specific day and even a specific hour in hour detail analysis. It will be also possible to choose between the different graph types.

1.4 Structure of this document

In sections 2 and 3 the background and the objectives of the thesis are presented. Afterwards, in section 4 the input data is described. Its processing and storage will follow through in section 5. Subsequently, the representation of the processed data will be shown and discussed in section 6. Several alternatives to this work will be proposed in section 7, allowing conclusions and possible upgrades to be developed in future works (section 8). Finally, a summary of the thesis will be provided (B).

2 Background

2.1 Psychiatric disorders review

Patients with severe mental illnesses have a higher probability of premature mortality. In particular, people with serious brain disorders die up to fifteen years earlier in comparison to the general population. Ruling out suicide and accidental death, coronary artery disease is a common cause of premature death for these patients [4].

The patients participating in this study suffer from schizophrenia, bipolar disorders or depression. Up to a 60% among these patients experience additional illnesses such as diabetes, respiratory or cardiovascular disease and hypertension[5]. These chronic diseases are associated with a sedentary behavior, so by analyzing the daily activity patterns of these patients we can increase their life expectancy. As an example, we know that depressed patients are physically sedentary[6]. When compared to the general population, their physical work capacity is reduced but their pulmonary function is normal. Therefore, this decrease of fitness level is caused by physical inactivity[7].

Evidence shows that the effects of exercise resemble those of psychotherapeutic interventions [8]. Due to its antidepressive effect[6], exercise alleviates depression and anxiety symptoms, improving overall their emotional well-being as the sense of being successful, in mastery and control will have positive affects in other fields of their lives[9]. These beneficial effects appear to be equal to those of meditation or relaxation[10] and don't carry the dangers or costs of drug therapy[11]. Furthermore, it helps to relieve some secondary symptoms of schizophrenia like low self-esteem and social withdrawal and it is also useful as a coping mechanism for auditory hallucinations, for instance[12].

An example of this type of physical therapy is walking; it can be performed either in supervised group walks or unsupervised home walking. It has become a really popular form of exercise for its simplicity, safety and low-cost[8]. It is a low intensity activity that can be easily performed by the elderly, the unfit and patients with mental illnesses[13].

On the other hand, excessive physical activity might lead to overexhaustion and overtraining, generating symptoms that simulate depression[10] as the patients are unable to comply to these high expectations.

In conclusion, by means of physical therapy, patients can improve their physical fitness, enhance their mental health by mitigating their symptoms and ultimately their social behavior[8]. Analyzing the daily activity patterns of the psychiatric patients, we can help to determine which patients need to increase their physical activity and which patients need to relax.

2.2 Types and characteristics of HAR sensors

In any HAR system, the first consideration is to select the type and number of sensors, including the position of the human body where they will be attached. The recognition of human activities has been mainly approached employing external sensors and those known as *wearables* [14].

When working with external sensors, they are usually placed in several points of interest located within a habilitated medical bay or in the patients' own home, where the recognition is performed. Although it can be considered an alternative with respect to *Internet of Things*[15], these external sensors systems depend completely on user interaction. For instance, whenever the user is out of the area controlled by these sensors, no data can be collected. Additionally, the installation and maintenance of this type of equipment can be quite costly.

The employment of cameras as external sensors can also be considered, given its suitability for interactive and security applications. However, video sequences entail a number of issues such as the privacy of the user, the ubiquity of video cameras and the complexity of video processing techniques (due to the massive data to be processed).

The main advantage of *wearables* is that they remain attached to the user. Given its small size, relatively low cost (in comparison to fixed sensors) and low energy consumption these devices have become quite popular in monitoring applications for rehabilitation of patients with brain disorders. Moreover, minimizing the number of sensors entails a reduction of the amount of data to process, its complexity and it is a more comfortable option [14].

The most broadly used sensors to recognize ambulation activities are triaxial accelerometers. As they are embedded in most of the currently used mobile phones, they are reasonably low-priced and require low power.

Maurer et al. [16] studied the relation between the sampling rate of the accelerometer (which lies between 10Hz[3] and 100Hz[17]) and the recognition accuracy only to find that there is no significant gain when recognizing ambulation activities above 20Hz. The sensor used for this project, SHIMMER3 [1], has a sampling frequency of 50Hz, although it is downsampled to 16Hz afterwards.

Another point of interest is the placement of the wearable sensors. He et al.[18] found that the optimal position on the body is inside the trousers pocket. Nevertheless, it depends on the goal of the application or the activities to be recognized; for example, placing the sensor in the wrist may not be the best option to recognize ambulation activities as any accidental movement might trigger incorrect predictions [14]. The position selected for the patients of this study is their lumbar area. We decided to use this position because most of the basic activities can be recognized accurately and patients find it comfortable.

2.3 Activity classification: Overview on HMM

A Hidden Markov Model (or HMM) is a highly capable statistical tool for modeling time series. In a HMM we consider that the observations are independent given the hidden process that generates the sequences[19], and any given state only depends on the previous state.

HMMs are modeled with three parameters[20]:

$$\lambda = (A, B, \pi)$$

- A, is the transition probability distribution matrix with dimensions $N \times N$, where N is the number of states in the model. The transition matrix stores the probability of state j (S_j) following state i (S_i), where S is the set of individual states denoted as $S = \{S_1,\ S_2,\ \ldots\ ,S_N\}$. In an homogeneous HMM, the transition matrix is time invariant [19].

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i))$$

- B, is the observation probabilities array with dimensions $N \times M$, where M is the number of observations. B stores the probability of the observation k being generated by state j (S_j). It is also time independent[19].

$$b_i(k) = P(x_t = V_k | q_t = S_i)$$

  where V is the set of observations denoted as $V = \{V_1,\ V_2,\ \ldots\ ,V_M\}$.

- π, is the vector of initial probabilities distribution with dimension N.

$$\pi_i = P(q_1 = S_i)$$

The Baum-Welch algorithm is used to learn the parameters of the model [21].

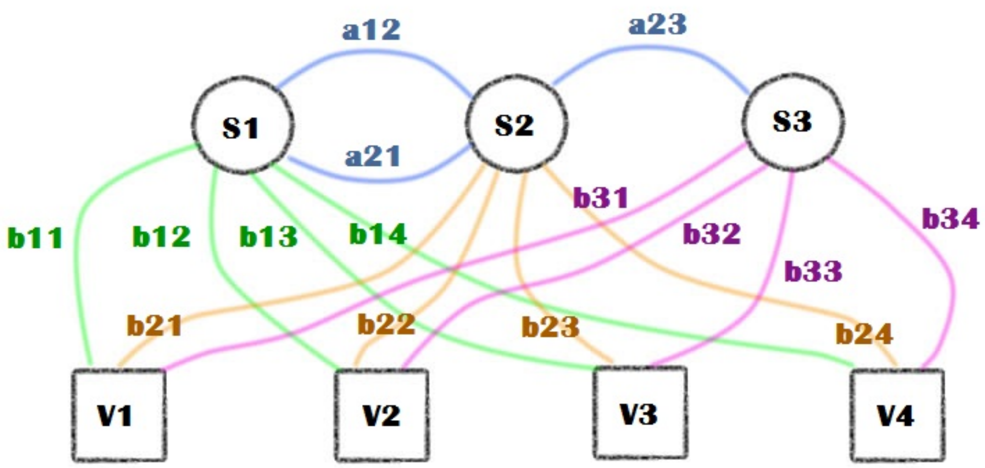


Figure 1: Example of an HMM model

In this work, each monitored activity corresponds to a different, independent, HMM[22] such as Han et al.[23] performed. The number of states of each HMM becomes a design parameter, in this work we employ five states per activity, following the model described in [20]. For every data sequence, we obtain the probabilities for each activity given the model using the Forward-Backward algorithm [21].

3 Objectives of the Thesis

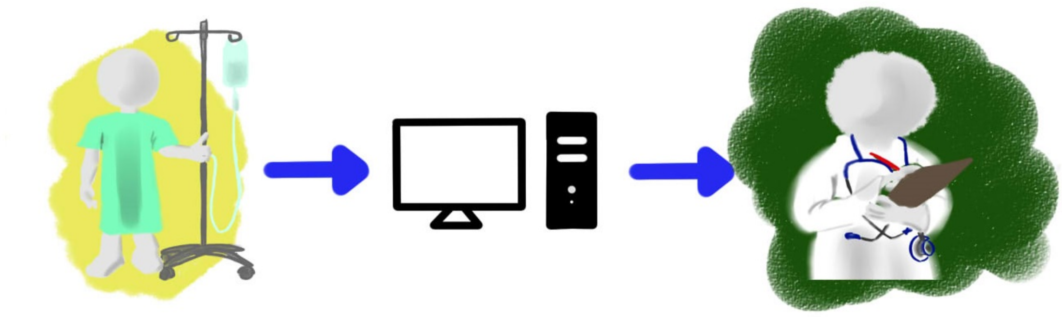


Figure 2: Data from the patient is processed to be later analyzed by the doctors

The objective of the thesis is to provide a daily activity patterns recognition system for a successful analysis of the patients behavior. For this purpose, we employ experimental data collected from psychiatric patients using wearable sensors. Furthermore, we provide useful data representations that help the doctors to interpret the behavior of their patients (Figure 2).

Due to the fact that these patients exhibit symptoms correlated to brain disorders (2.1), physical therapy is part of their treatment. Moreover, patients are required by doctors to follow an activity schedule (as it can seen in Figure 3) where this physical therapy, as well as group psychotherapy, is included. In order to keep an eye on their progress, doctors are interested in keeping track of their daily activity patterns to verify it matches the schedule.

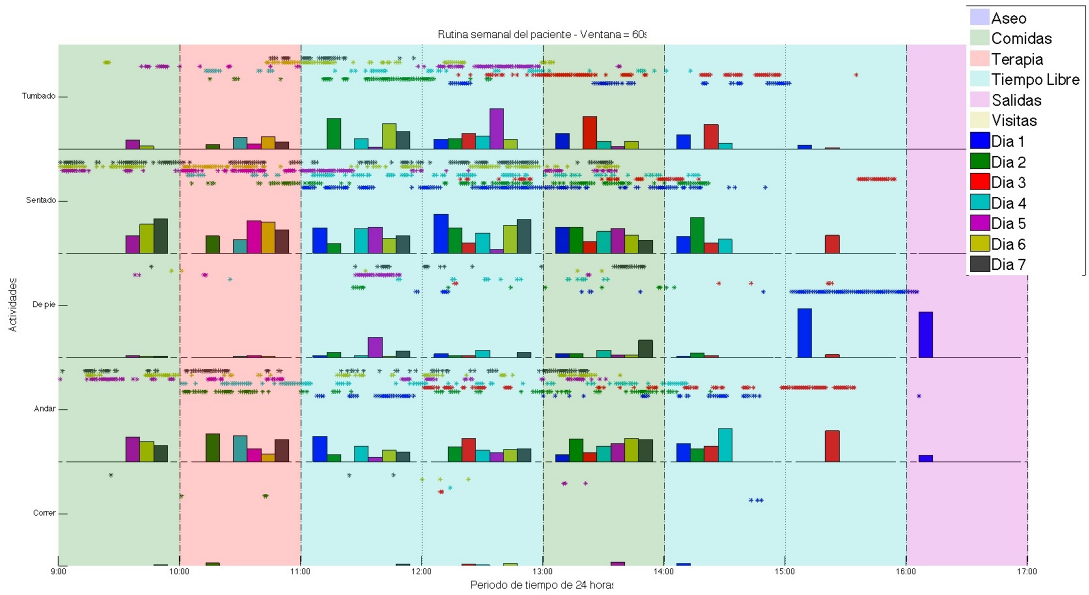


Figure 3: Example from previous stages of the project.

The patient data provided by the doctors must be read and accessed in a correct manner, following the specific format given by the previous steps of the project.

The sensor taking these samples (Figure 4) is being carried by the patient at all times from the moment it is turned on. This sensor is placed on the patient's lumbar area in order to be as unobtrusive as possible.

(a) Side and front view

(b) Side view with strap

Figura 4: Sensor used: SHIMMER3©

The data sampled in each time instant, that is, the action being carried out by the patient in that very time instant, is interpreted and transformed into five probabilities that correspond to the five activities considered in this project (section 4.2) . In other words, for each time instant there are five activity probabilities.

In order to achieve such interpretation, we must assure that the proposed method works for an undetermined number of patients monitored and for an undetermined number of days. This means that data storage must provide enough capacity for the given amount of information.

In a production environment all input data should be checked to assure that certain conditions are met. First, all samples belonging to the same recording must be filled. As they are posterior probabilities of the activities being performed, all the observations must follow the fundamental probability theorem: Each probability value must be between zero and one, and the sum of all probability values for each instant must be equal to one. Additionally, these values must always be numerical. Textual characters or special symbols (like &, %, _) will be discarded. As these samples are the result of prior sensor data processing [22][20], the situations described above (ideally) should not happen.

The input data belonging to the patient needs to be accessed and classified to enable its posterior analysis and representation. For this purpose, input data will be stored in a classified and identifiable cell unit in matlab (section 5.2).

Input data will be processed and analyzed according to the variables that are of interest to the doctors. A graphical user interface is proposed to ensure such performance and flexibility, allowing to introduce the data selection parameters.

For the purpose of this project, the end user will be able to select: a specific patient and a specific day for analysis and representation. As a result, all recordings of this patient for that day will be displayed.

A graphical display of selected data is provided to the user. Three main representations are offered: day intensity, day evolution and hour details.

4 Description of the Daily Activities Database

A total of 25 patients of the psychiatric wing of Fundación Jiménez Díaz have worn the sensors and provided enough data to carry out this study. Each patient was assigned a random number with four to six digits in order to keep anonymity (Figure 5). This number will be further referred to as *patient_ID* and it's assigned as the patient's folder name.

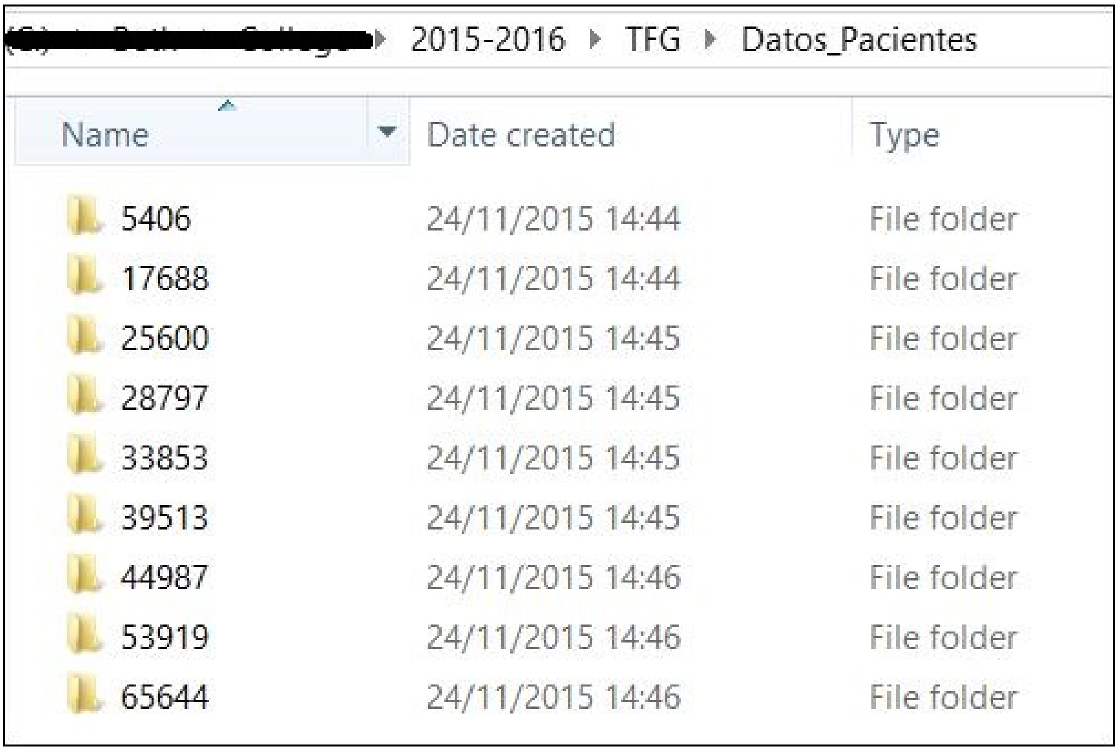


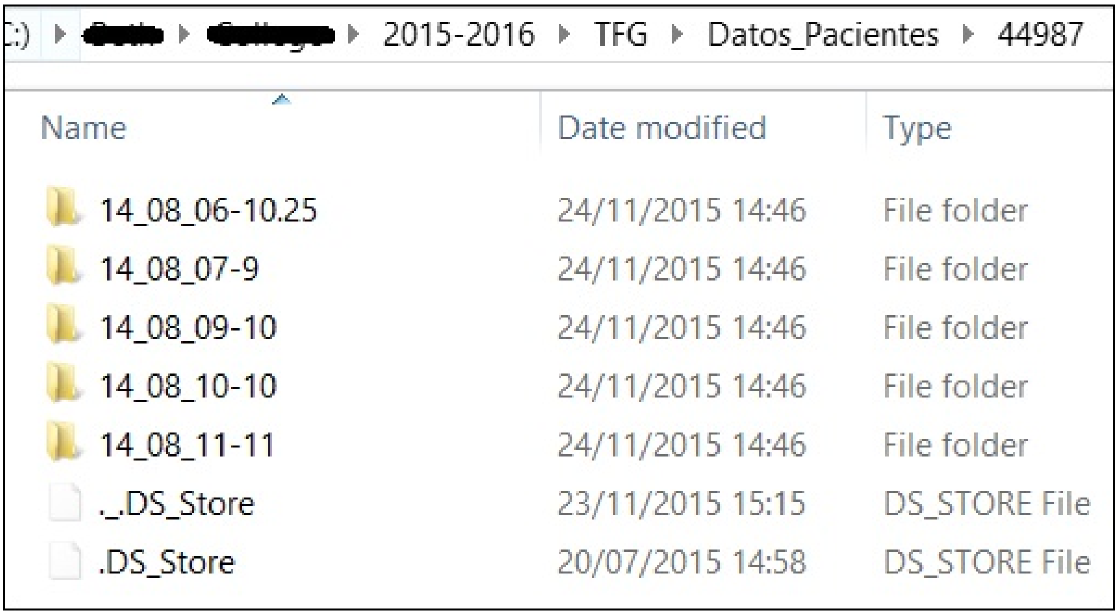
Figure 5: Example of a list of patients.

4.1 Patient folder format

Within each patient's folder, it is provided a subfolder corresponding to each day recorded and the subfolder name establishes the date and starting time of each session (Figure 6). The format used is YY_MM_DD-Time:

- **YY,** stands for the year.
- **MM,** stands for the month.
- **DD,** stands for the day.
- **Time:** recording starting time in decimal format (hours and cents of an hour). Timing goes from 0h-24h.

For instance, 14_08_06-10.25 corresponds to August 6th of 2014 at 10:15AM.



:) ▸ [redacted] ▸ [redacted] ▸ 2015-2016 ▸ TFG ▸ Datos_Pacientes ▸ 44987

Name	Date modified	Type
14_08_06-10.25	24/11/2015 14:46	File folder
14_08_07-9	24/11/2015 14:46	File folder
14_08_09-10	24/11/2015 14:46	File folder
14_08_10-10	24/11/2015 14:46	File folder
14_08_11-11	24/11/2015 14:46	File folder
._.DS_Store	23/11/2015 15:15	DS_STORE File
.DS_Store	20/07/2015 14:58	DS_STORE File

Figure 6: List of recordings of patient 44987.

Not all day recordings state their starting time and it has to be taken into account during data processing.

In each day subfolder, pacient activity data is contained in a single ".mat" file for that day (Figure 7). This type of files are matlab data files. File format is YY_MM_DD_patient_ID_probs.mat.

[redacted] ▸ 2015-2016 ▸ TFG ▸ Datos_Pacientes ▸ 44987 ▸ 14_08_06-10.25

Name	Date modified	Type	Size
14_08_06_44987_probs.mat	24/02/2015 17:03	MATLAB Data	24,393 KB

Figure 7: Data file of patient 44987 for August 6th of 2014.

4.2 Patient data format

The data file for each day contains a data matrix in which the probabilities of a patient performing one of the five activities considered in a time instant are listed. This data matrix is called *activities_probabilities*, its dimensions are $A \times N$, being A the number of activities and N the number of samples (time instants).

The values for each sample, as a probability, is between 0 and 1. The sampling frequency is 16Hz, providing a sampling period equal to 0.0625s.

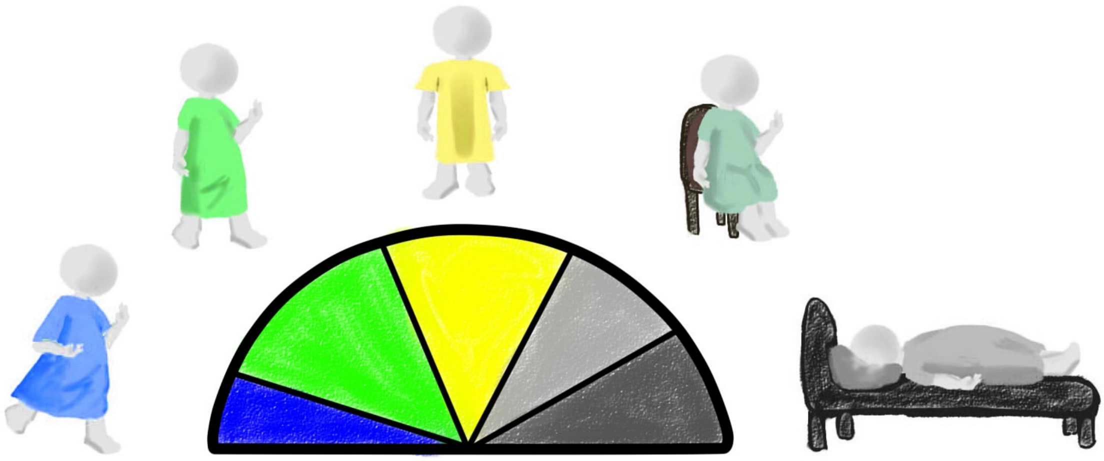


Figure 8: Activities: running, walking, standing, sitting down and lying down.

For this project, the registered activities (Figure 8) are five:

- **1st row,** running. Energetic activity, running or walking fast.
- **2nd row,** walking. Device is upright and detecting rhythmic movements consistent with walking.
- **3rd row,** standing. Patient is at upright position, small movements give away the activity (shifting weight from one leg to another).
- **4th row,** sitting. Patient is at upright position and there is no movement.
- **5th row,** lying down. Device is at horizontal position, so patient must be lying mostly static and not moving much.

5 Data Processing

5.1 Program parameters

Data processing requires some basic information about how the data samples have been recorded and we must define some basic parameters to fix the program behavior and information display.

Sampling frequency of the sensor, number of activities to consider, starting time, ending time and time step are considered constants declared within the code:

$hstart$ = 8; (8:00, standard start time)
$hend$ = 24; (24:00, standard end time)
$fsampling$ = 16; (16Hz)
$time\_step$ = 0.25; (15mins, fraction of hour)
$states$ = 5; (5 activities to be considered)

Let us consider the following variables:
patient_ID, this is the number used to characterize the patient data.
input_date, this is the date of the file to be analyzed.
input_hour, this is the hour selected for detailed analysis.

5.2 Data storage

Initially, the application will need to find the original patient data. It will ask the user to select the folder in which the data for all patients is contained (Figure 9). As a result, the path for the directory containing all input data is obtained. This path is necessary for input data retrieval.

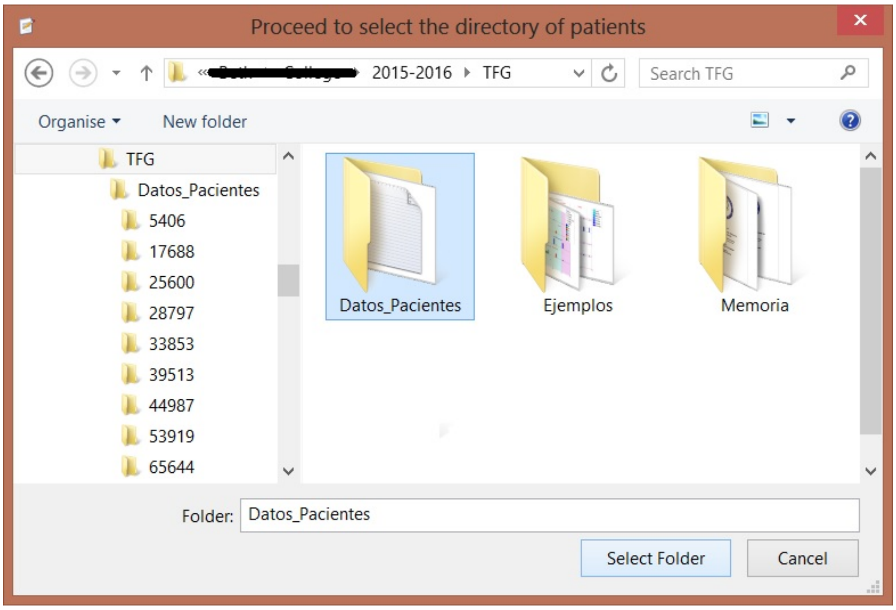


Figure 9: Capture of the initialization of the graphical user interface

The user will then introduce the *patient_ID*. This *patient_ID* is concatenated to the previous data path, obtaining as a result the full path corresponding to the specified patient. The user must also specify the date of interest. With these parameters the program will select the proper daily activity pattern data file and it will be subsequently loaded into a matlab cell unit.

The time frame with the starting time and the ending time, and also a time interval for data aggregation are constants within the program. Due to the patients' schedule, the proposed times for this time frame are eight o'clock and midnight with a time step of fifteen minutes.

For example: A doctor is interested in accessing recorded data of the patient *patient_ID* on an specific date or hour. In order to do so, the doctor needs to input the *patient_ID* and the desired date and select the hour he wants to inspect. The program will divide the time range by the corresponding time step indicated to obtain the number of columns that the resulting cell will have (because it is only a specific day, this cell will have only one row).

In this way, the input recording of the patient will be ordered and stored just as the doctor has requested.

Figure 10: Each subcell of the cell corresponds to a piece of input data

On this project we have decided to employ cell units to store patient data instead of matrices due to its simplicity and ordered manner. It would not be appropriate to use matrices because the input data samples would create matrices of massive dimensions and the access to such structures would slow down the whole process. Also, keeping track of the proper index to access the elements in the matrix would be quite troublesome.

Additionally, cell units allow the storage of multiple data types (double, char, vectors, matrices...) of different dimensions simultaneously. That is, a subcell can contain a string of 72 characters and the following subcell a vector of only two elements. Matrices cannot compete with this advantage.

For this project, we are interested in storing matrices as the elements of the cell. Each matrix stores the input samples corresponding to the date selected.

Also, if for a determined time instant there is no input data, the cell unit elements equivalent to that time instant will be null. There would be no additional matrices full of zeros consuming memory and resources; a characteristic truly attractive for the visualization of the recorded input data.

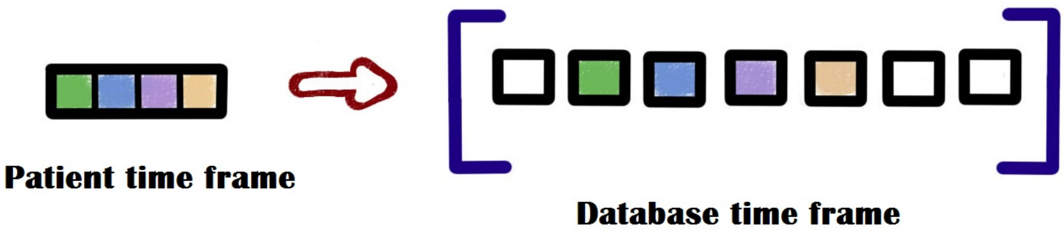


Figure 11: Cell filled with the corresponding daily activity pattern. Empty subcells have null value.

Another point to take into account is the volume of data processed. If we consider a daily activity pattern file starting at eight o'clock and lasting until midnight, we obtain a total time range of sixteen hours. As the sampling frequency is 16Hz the sampling period is 0.00625s.

For a minute, 960 samples have been taken.

For fifteen minutes, 14400 samples.

For an hour, 57600 samples.

For sixteen hours, a total of 921600 samples have been gathered. On top of that, for each sample taken there is a probability associated to each possible state.

That means a total of 4608000 probabilities for a single day for a single patient.

The cell storing these probabilities would have dimensions 1x64; one row because it is only one day, and sixty-four as the result of dividing the total number of minutes (16h = 960 minutes) into subcells of fifteen minutes. Each subcell will host a matrix storing input samples for a time step of fifteen minutes.

With these parameters a day data file is around 24MB in size. For this project we have available around 300 days belonging to several patients for a total worth of near 6GB of data.

5.3 Data classification

A function called "cellSamples.m" will return the cell corresponding to the input patient data, for the specified date and ordered according to its starting time. It's within this function where the original daily activity pattern file is accessed and loaded.

The input needed for this function are *patient_ID*, *input_date*, *hstart*, *hend*, *time_step* and *fsampling*.

As stated in the previous part (5.2), in order to do so, the full path for the patient data is completed with the contribution of *patient_ID*. Each folder is checked and the one corresponding to *input_date* has its daily activity pattern file loaded into a cell created at the beginning of the function. To achieve the correct order of the samples, we need to be very careful with the starting time of the input sample. That is, the first column of the cell that will be filled with data.

The length of time that each individual has been wearing the sensor, as well as the number of days recorded, varies with each patient so we have considered a maximum time range from eight o'clock in the morning until midnight. That is, a maximum of sixteen hours per day.

The starting time of the daily activity pattern is the last part of the name of the patient folder (4.1), therefore these characters will be read and transformed into a numerical value. Nevertheless, not every folder follows this pattern. If it were the case, data belonging to such recording will be treated as starting at eight o'clock in the morning. On top of that, the values corresponding to the starting time will be always given in decimal format. Basically, the starting hour will always be in quarters: either o'clock, quarter past, half past or quarter to.

As a result, a vector containing all possible values from eight o'clock in the morning until midnight in steps of fifteen minutes will be used to determine the position of the starting hour read (or assumed). To sum up, this vector will return the value of the first column of the cell that will be filled with data. The function that acts as translator between an hour and its column position is called "hour2index.m".

The resulting cell, with dimensions 1×64, is returned. Let us call this result *patient_cell*.

5.4 Data selection

The function called "cellMean.m" will return the matrix containing the mean of every time step of the input patient data. Within this function, daily activity pattern files are simplified enough to obtain a representation in matlab.

The input needed for this function are *states* and the resulting cell from function "cellSamples.m".

Each subcell corresponds to fifteen minutes of the time range, containing a maximum of 14400 samples. In order to reduce the workload, function "cellMean.m" computes the mean of these 14400 samples for each state.

A matrix containing all sample means for each subcell and dimensions $states \times Ncolumns$ is returned. In this work, its dimensions are 5×64. Let us call this resulting matrix *matrix_day*.

The last of these data processing functions is "cellMode.m", which returns the vector containing the mode (the activity more repeated as highest probability) of every minute of the input patient data. Within this function, daily activity pattern files are simplified enough to obtain a representation in matlab.

The input needed for this function are *input_hour*, *hstart*, *hend*, *time_step* and *fsampling* and the resulting cell from function "cellSamples.m"

For this function we are interested in obtaining the mode of each minute in the subcell selected by *input_hour*, and the next three (four quarters). Basically, we are selecting a whole hour and computing the mode corresponding to each minute.

By means of translating *input_hour* into its column position with function "hour2index.m", the subcell it refers to and the following three will be selected.

A minute consists of 960 samples and its mode represents the activity most frequently repeated with highest probability in all 960 samples, which is later stored in a vector that will be returned.

The resulting vector, with dimensions 1×60 (one result per minute), is returned. Let us call this result *matrix_hour*.

6 Graphical User Interface

6.1 Application input form

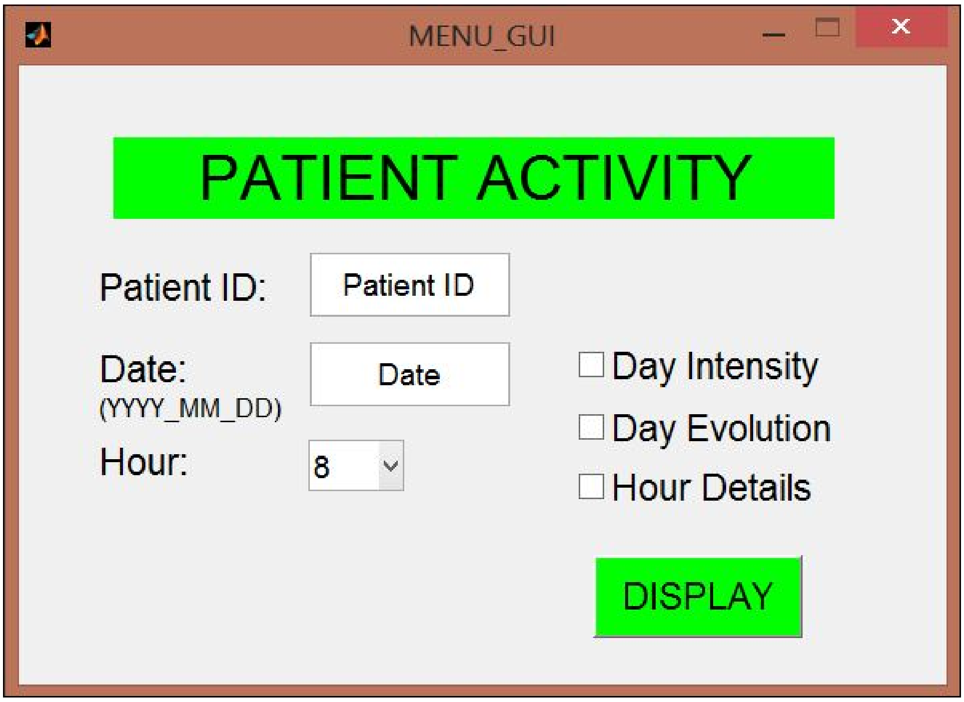


Figure 12: Capture of the graphical user interface.

Once the proper directory for the patients is selected and the graphical user interface is loaded, the user needs to introduce the following parameters:

- **Patient_ID,** random number with four to six digits assigned to the patient to ensure anonymity. Example: 5406, 44987, 105680...
- **Date,** complete date of the day of interest. Needs to follow the format YYYY_MM_DD to access the right folder.
- **Hour,** marks hour of interest. The values of the hour go from eight o'clock in the morning until eleven o'clock in the night, in steps of one hour.
- **Day Intensity,** check box with boolean value. If it is checked, has value equal to one and the result will provide a temporal plot of each activity performed by the patient during the day selected.
- **Day Evolution,** check box with boolean value. If it is checked, has value equal to one and the result will provide an area plot of the activities performed by the patient during the day selected.
- **Hour Details,** check box with boolean value. If it is checked, has value equal to one and the result will provide an area plot of the activity level performed by the patient during the selected *hour*.
- **DISPLAY,** button to start the processing of the input data of *patient_ID*.

6.2 Data visualization

Let us consider the following example:

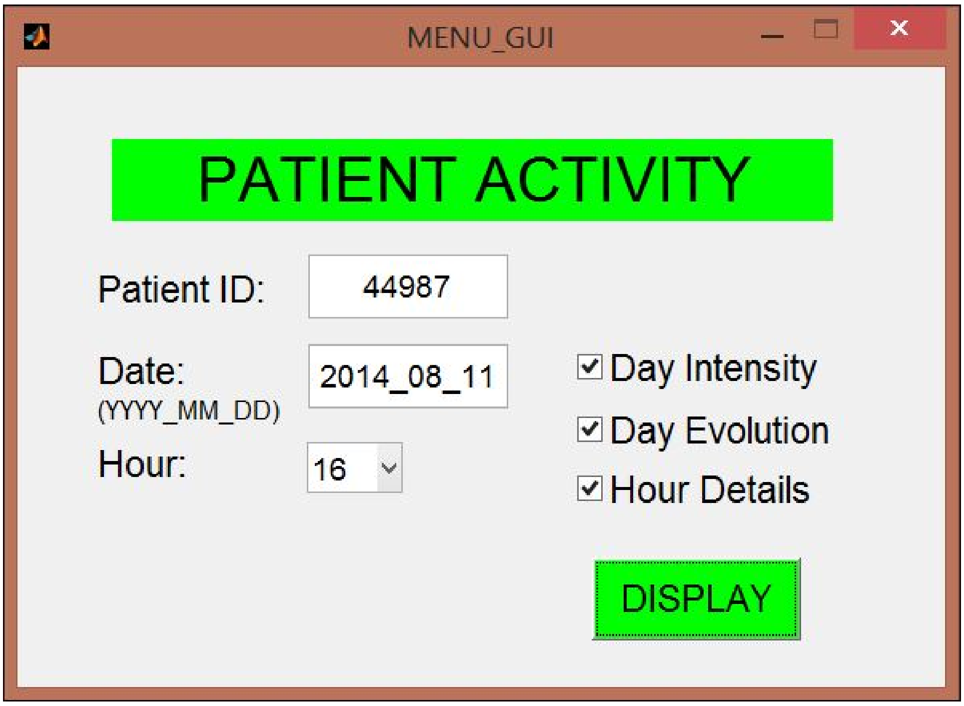


Figure 13: Example of the input data in the GUI.

Once the button DISPLAY is pressed, the program gets the values selected for the following variables: *patient_ID*, *input_date*, *input_hour*, *day_INTENSplot*, *day_EVOLplot* and *hour_plot*. These variables are a necessary input to the data processing functions (5.3)(5.4).

- *patient_ID* = 44987;
- *input_date* = 2014_08_11;
- *input_hour* = 16;
- *day_INTENSplot* = 1;
- *day_EVOLplot* = 1;
- *hour_plot* = 1;

There are three available representations of the daily activity pattern file:

- **Day Intensity**, as shown in Figure 14, represents the intensity of the activities performed by the patient. It's a temporal plot of the sampled probability means of the activities, collected in *matrix_day*. As the hue gets more intense, the mean for the activity increases.

  The x-axis represents the time of the day from eight o'clock until midnight in steps of fifteen minutes (only sharp hours are labeled for reference). The y-axis represents the activities performed by the patient.

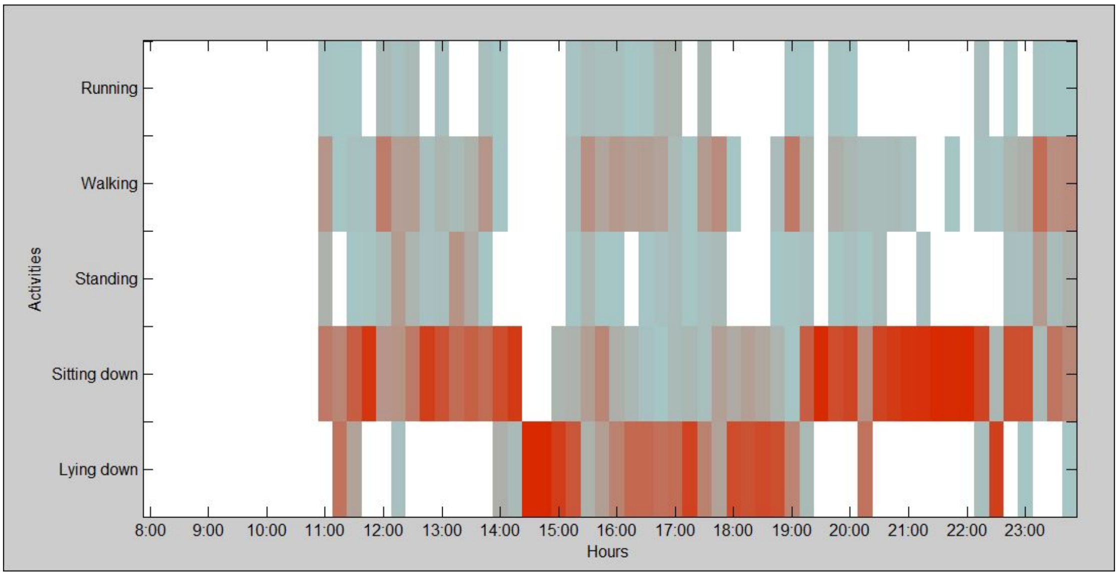


Figure 14: Day intensity graph for the example proposed.

Patient 44987 on August 11th of 2014, with recording starting at eleven o'clock in the morning, was mostly seated and lying down.
This graph represents an individual with a really low level of activity, as it can be seen in the segments between [11:00-14:30], [14:30-19:15] and [19:15-24:00] the highest probabilities are a darker color. Although, the patient seems to have been moving around some time around the day.

An earlier version of the graph obtained for day intensity is shown in Figure 15, with the y-axis listing the available days of the patient and the x-axis displaying the mode of samples (instead of the sampled mean) obtained per minute of the day. As it didn't suggest any clear interpretation of the recordings of the patient, this idea was discarded and we decided to focus on the representation of activities on a single day.

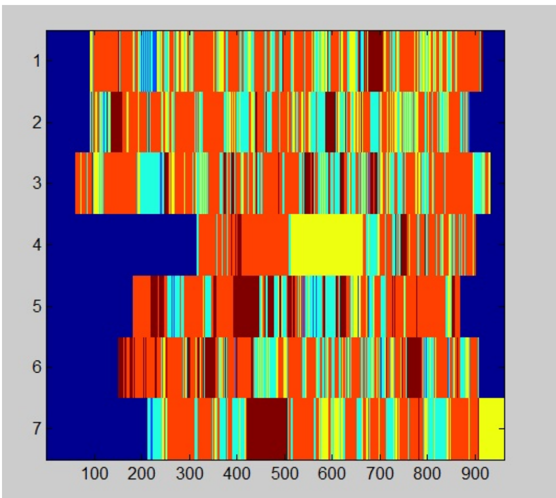


Figure 15: Day intensity, previous version for patient 105680

- **Day Evolution**, as shown in Figure 16, represents the evolution of the activities performed by the patient along the day. It's an area plot of the sampled probability means of the activities, collected in *matrix_day*. A darker hue stands for the more energetic activities, being "running" (high activity) the maximum with the darkest color. In the same way, a lighter hue suggests a minimum activity.

  The x-axis represents the time of the day from eight o'clock until midnight in steps of fifteen minutes (labels are only shown for full hours). The y-axis represents the accumulated probability of the activities performed by the patient, with high activity at the bottom and adding lower and lower on top of it. They will add up to one (100%), as should be expected.

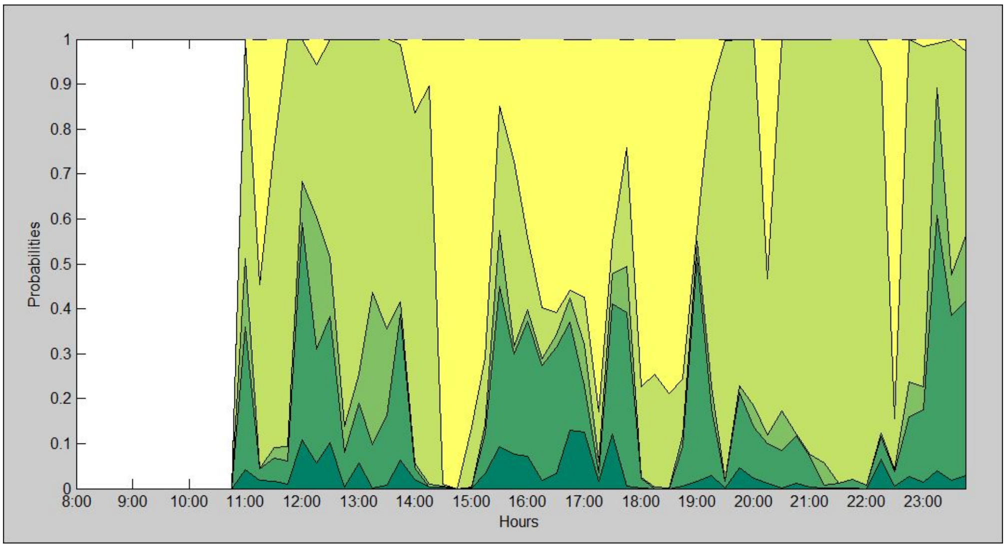


Figure 16: Day evolution graph for the example proposed.

Patient 44987 on August 11th of 2014, with recording starting at eleven o'clock in the morning, was mostly seated and lying down.

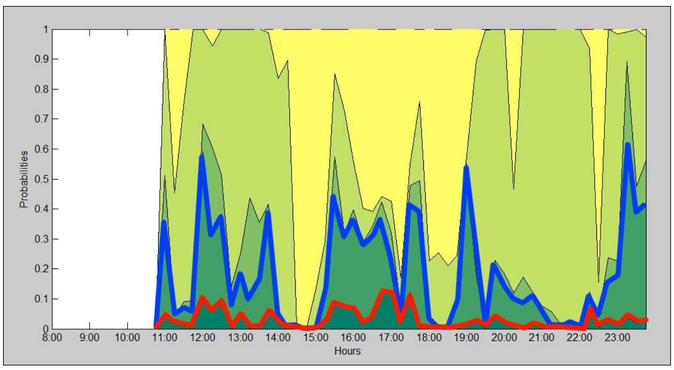


Figure 17: Activity lines in day evolution graph for the example proposed.

In this graph, one can discern the lines corresponding to the probability of each activity. In Figure 17 the marked lines are the ones that represent high activity; red stands for running or fast movements and blue stands for walking.

- **Hour Details**, as shown in Figure 18, represents the activities performed by the patient. It's an area plot of the modes of the activities, collected in *matrix_hour*. The x-axis represents the minutes of the selected hour, from one until sixty in steps of one minute. The y-axis represents the activities performed by the patient.

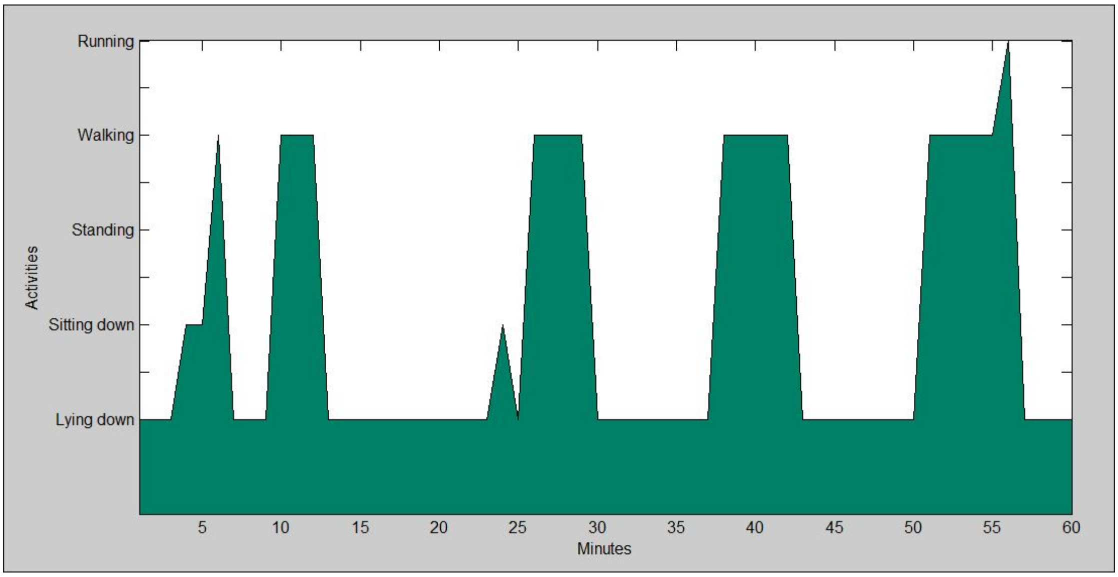


Figure 18: Hour detailed graph for the example proposed.

Patient 44987 on August 11th of 2014, between four o'clock and five o'clock in the afternoon, was mostly walking and lying down as it can be seen by the alternation of states.

An earlier version of the graph obtained for hour detailed is shown Figure 19, with the y-axis stating the list of activities in the order proposed by previous stages of the project (1 - running, 2 - walking, 3 - standing, 4 - sitting down and 5 - lying down) but as it suggested incorrectly that an increase in the graph was an increase of activity, the order was reversed.

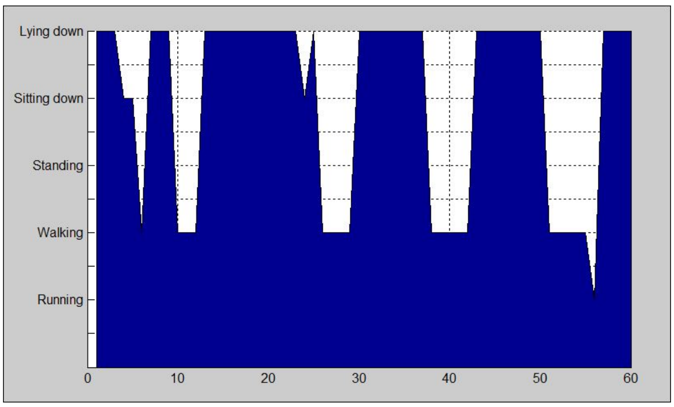


Figure 19: Hour detailed graph before editing

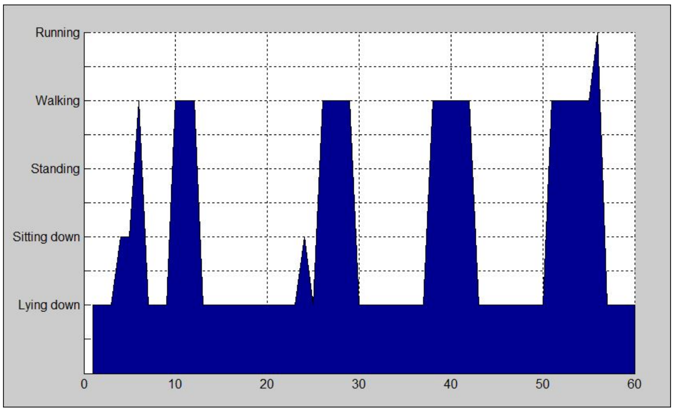


Figure 20: Hour detailed graph after editing

In Figure 20 the order of activities has been reversed (1 - lying down, 2 - sitting down, 3 - standing, 4 - walking and 5 - running) for a more helpful approach and easier visualization.

6.2.1 Graph interpretation

In the following pages,real graphs corresponding to two different days of patient 44987 are shown.

Figures 21 and 22 are the day intensity and day evolution graphs for August 6th of 2014. On this day in Figure 21 (day intensity graph) we can see better the resting activities as they are continuous in time, so they are more apparent, while in Figure 22 we can see better the activity periods as they add up from the bottom. In this last graph, we can see clearly two activity periods in the morning and afternoon, with a resting period during [14:00-15:30].

Figures 23 and 24 are the day intensity and day evolution graphs for August 10th of 2014. On this day it can be observed that the patient was intermittently active during the day with a resting period around [13:00] and [16:00-21:15] until the end of the recording.

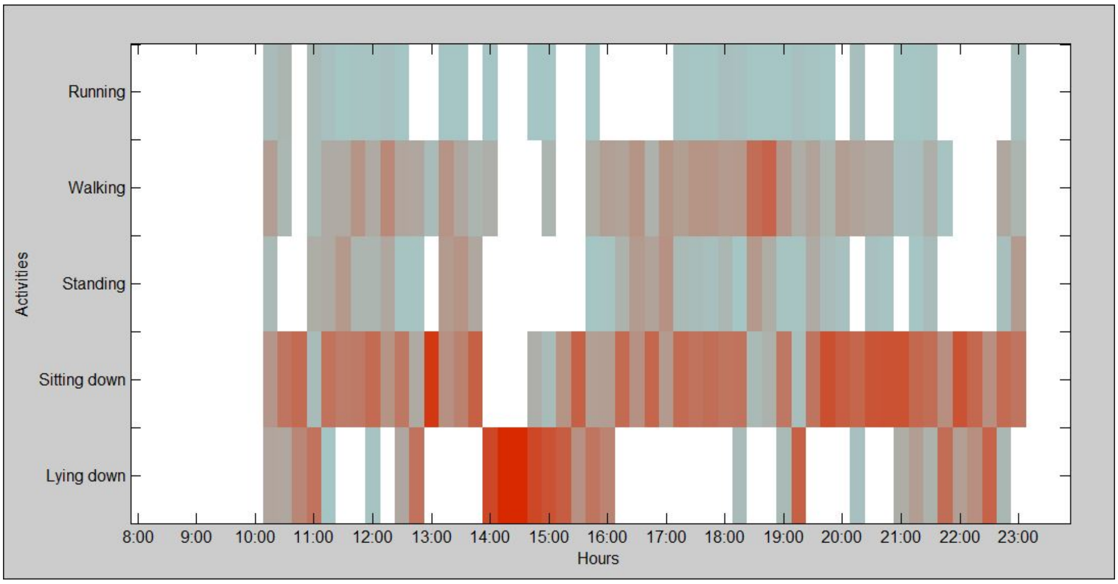


Figure 21: Patient 44987: Day intensity graph (2014_08_06).

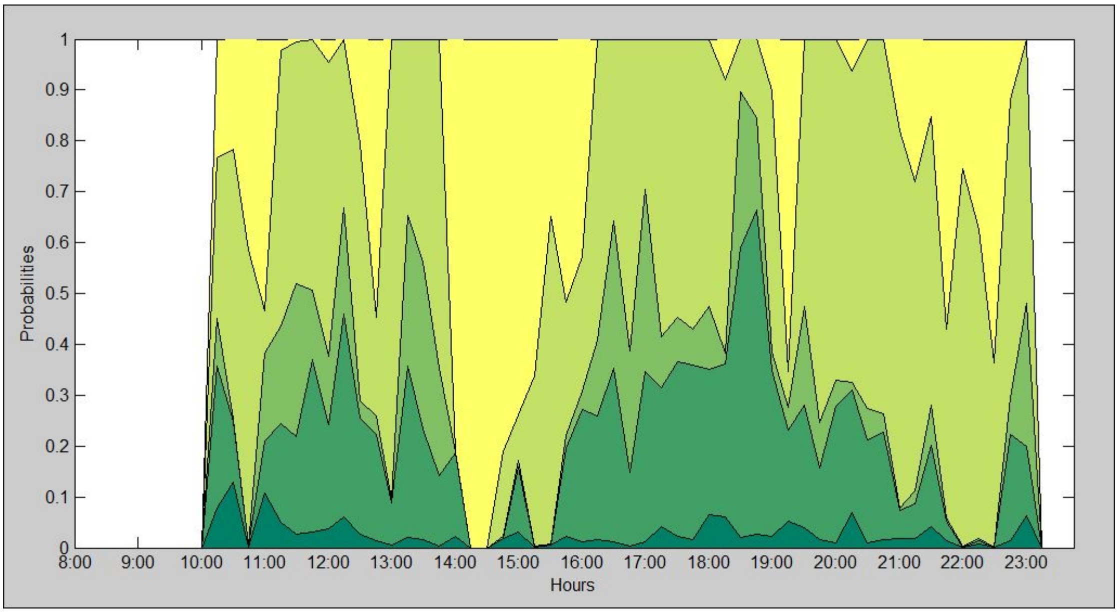


Figure 22: Patient 44987: Day evolution graph (2014_08_06).

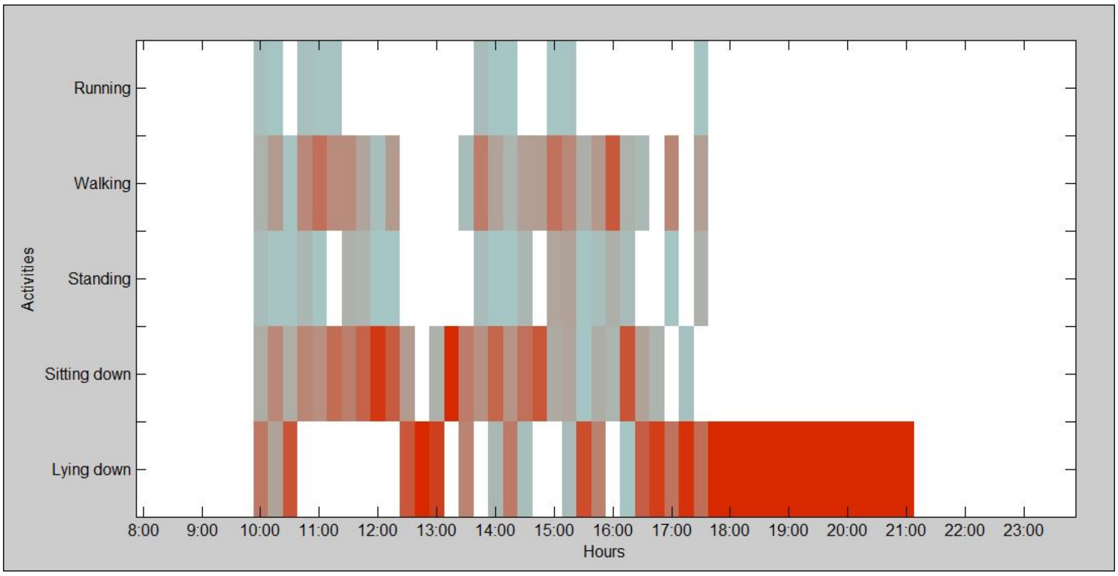


Figure 23: Patient 44987: Day intensity graph (2014_08_10).

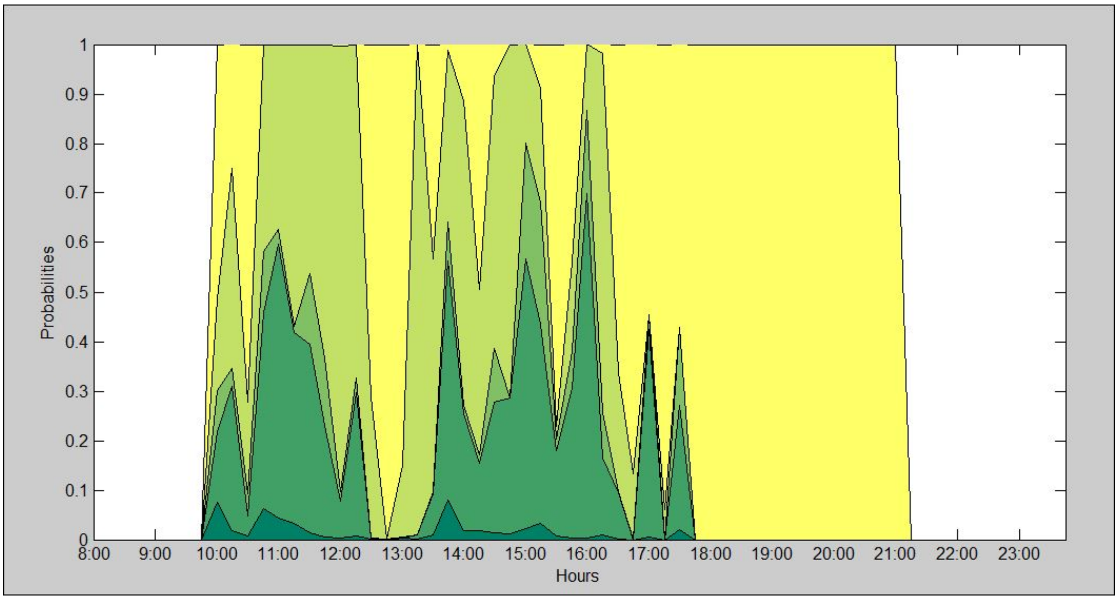


Figure 24: Patient 44987: Day evolution graph (2014_08_10).

7 Alternatives

Let us discuss some possible alternatives to the work done:

- Data processing could be performed in other environment besides matlab. Number-crunching programs such as R or even Excel are an option to consider. Also, mathematical libraries in compiled languages like C could be used.

- Once data has been processed for presentation, graphs can be generated with specialized libraries for more appealing displays for the end user (like plot.ly within matlab). Other graphical libraries and compiled languages could be used. Even Excel could be also considered for display.

- By using a website, the end user interface would be easier to create and modify. All data processing and graphic images would be processed by the web server providing non-local access to the application. This processing in the background would allow independent programs for the different tasks, like file handling, data processing and visualization. Of course, this rises the concern for patient data protection. Access control system must be implemented.

8 Future Work

In this thesis, we have overcome the task of representing the set of data, daily activity patterns, obtained in previous stages of the project and even have suggested a mock graphical user interface for the end user.

The next step of the project is to develop the final application that will be used by medical personnel. That is, to program such application in Android or similar programming languages to implement all work done in this project. This final application should be accessible from either computer, tablet or smartphone, in order to check on patients easily from anywhere should a crisis or urgent need arise.

However, a fundamental step is to confirm whether the proposed graphical representations are useful to doctors, with the aim of establishing a level of adequacy of these representations of patient behavior. After all, that is the main goal.

Doctors may suggest additional characteristics or highlight certain features that appeal to them.

Suggestions for new developments could be:

- Gathering subject data with smartphones. In less critical patients like those in dieting or rehabilitation, using medical devices could be unnecessarily costly or cumbersome. Smartphones do have accelerometers and position sensors to provide basic activity monitoring, also differentiating if the phone is being used or the patient is moving.

- It could be interesting to check on multiple subjects at once to compare their behaviors in patients' group therapy, for instance. In order to do so, the end user should be able to introduce the *patient_ID* for more than one patient. This could be applied to athletes training, to monitor their progress.

- Reviewing longer periods of time to check the progress of the patients. However, for a adequate visualization, data should be more continuous for consistency. As the data presently available is from non-consecutive days, this proposal would create today a gap between samples.

- Make available a written patient report for the day, in text and graphics format. It would be also appealing to have such report for a week, as these reports would be quite useful for other medical applications such as dieting, keeping track of athletes with a particular training or in a rehabilitation process.

- Patient feedback would also be a great option to consider. Via messages on the mobile app on their smartwatches or smartphones, patients could receive the aforementioned reports with their progress or statistics of the day. This feedback could be modulated by some parameters introduced by the doctors, in order to motivate them (patients) or limit them (athletes).

  An example of these messages could be "You have walked a total of __ steps! Well done, you only need __ to reach your goal!".

A Annex 1: Project Scheduling and Budget

A.1 Project scheduling

The following table shows the total time needed for the fulfillment of the project. Given that this project is a Bachelor Thesis, the total number of hours required must be equal to 300 hours.

There are four main sections:

- **Documentation.** Investigation and inspection of the samples obtained in previous stages of the project and research about the basics of human activity recognition, as well as reading papers regarding previous work.
- **Matlab coding.** This is the core of the project. Reading input data, storage of said data, selection and representation. It also includes all programming of the graphical user interface.
- **Results.** The obtained graphs and results are examined and tested to check their consistency. Modifications of the previous matlab coding are included here.
- **Report.** Writing this very report requires certain knowledge of programming in LaTeX. Double-checking for errors or grammatical mistakes and the design of the presentation for the thesis defense.

All the hours needed with the supervisor of the project to check the progress of the project itself and solving doubts about it are also included.

Task		Duration	ID
Documentation			
	Literature about HAR	20 hours	A
	Understanding the input samples	10 hours	B
Matlab coding			
	Development of scripts	60 hours	C
	Matlab library inquiries	15 hours	D
	Graphical user interface development	20 hours	E
Results			
	Tests	30 hours	F
	Optimization	10 hours	G
Report			
	Learning LaTeX[24]	15 hours	H
	Document format	5 hours	I
	Report writing	70 hours	J
	Edition and correction	18 hours	K
	Presentation	15 hours	L
Tutoring		12 hours	M
TOTAL		300 hours	

Table 1: Project scheduling

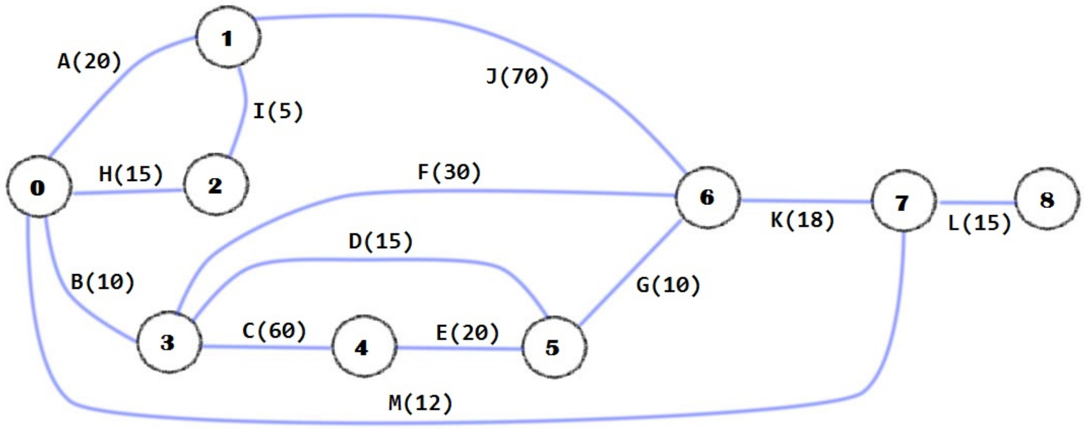


Figure 25: PERT diagram for this project.

A.2 Project budget

In order to properly assess the total cost of the project, it is necessary to distinguish between personnel costs and material costs.

- **Personnel costs**: Entails the salary of all people participating in the project.

  According to the Colegio Oficial de Ingenieros de Telecomunicación[25], the average salary of an electrical engineer participating on a project is of 60€per hour. As stated in A.1, the total number of hours for this project is 300 hours and the length of this project has been of five months.

PERSONNEL COSTS	Rate [€/hour]	Monthly cost [€]
Engineer's salary	60.00	3600.00
Project supervisor's salary	150.00	360.00
Sub-TOTAL	210.00	3960.00
TOTAL COST IN 5 MONTHS		19800.00€

Table 2: Total amount of personnel costs.

- **Material costs**: Include software license, computer maintenance and minor expenses such as printing research the papers essential for the project.

MATERIAL COSTS	Cost [€]	Monthly cost [€]
MATLAB Student License[26]	69.00	5.75
Laptop	1800.00	50.00
Other costs	50.00	10.00
Sub-TOTAL	1919.00	65.75
TOTAL COST IN 5 MONTHS		328.75€

Table 3: Total amount of material costs.

- **Total costs**: The total budget necessary to cover the expenses for the project is the sum of the total personnel costs and the total material costs, plus a 21% for taxes.

TOTAL COST	Cost [€]	21% IVA
PERSONNEL COSTS	19800.00	4158.00
MATERIAL COSTS	328.75	69.04
Sub-TOTAL	20128.75	4227.04
TOTAL COST (w/IVA		24355.79€

Table 4: Total budget for the project

B Annex 2: Summary

B.1 Introduction

Psychiatric patients suffering from disorders like depression or schizophrenia develop activity patterns that show their psychological state. The monitoring of those activity patterns is a valuable tool that doctors can employ to evaluate the health conditions of their patients and decide their treatment. This monitoring could be achieved through different types of devices that should be unobtrusive and easy to wear for the patients and easy to handle by the medical staff.

Today, the rise of wearable devices such as *smartwatches* (Pebble Time, Moto 360, Apple Watch...) and other attachable gadgets (fitbit, Samsung Gear Fit...) provides an excellent opportunity to introduce sensors in medical applications. Patients can be monitored using *wearable* sensors to proportionate accurate daily activities data whether they are in a restricted medical environment or in ambulatory conditions. In people like psychiatric patients suffering from bipolar disorders, depression or schizophrenia offers the possibility of carrying out rehabilitation from home and can be remotely monitored without the need of being hospitalized.

Moreover, public acceptance of other types of wearable devices such as heart rate monitors and fitness trackers indulges this train of thought. As they are no longer only associated to medical purposes, like Holter monitors or insulin pumps, anyone is able to wear a sensor without psychological prejudices or physical disturbances.

B.2 Background

B.2.1 Psychiatric disorders review

The patients participating in this study suffer from schizophrenia, bipolar disorders or depression. These patients experience additional illnesses such as diabetes, respiratory or cardiovascular disease and hypertension[5], which are associated with a sedentary behavior. As an example, we know that depressed patients depressed patients are physically sedentary[6].

The effects of exercise resemble those of psychotherapeutic interventions [8], alleviating symptoms and improving overall their emotional well-being. These beneficial effects appear to be equal to those of meditation or relaxation[10] and don't carry the dangers or costs of drug therapy[11].

It helps to relieve some secondary symptoms of schizophrenia like low self-esteem and social withdrawal[12].

In conclusion, by means of physical therapy, patients can improve their physical fitness, enhance their mental health by mitigating their symptoms and ultimately their social behavior[8]. Analyzing the daily activity patterns of the psychiatric patients, we can help to determine which patients need to increase their physical activity and which patients need to relax.

B.2.2 HAR sensors

First consideration for Human Activity Recognition is to select the type and number of sensors, including the position of the human body where they will be attached.

The recognition of human activities has been mainly approached employing external sensors and those known as *wearables* [14]. These sensors remain attached to the user and are characterized by its small size, low cost and low energy consumption. By reducing the number of sensors, the amount of data and its complexity are reduced as well [14].

Most broadly used sensors to recognize ambulation activities are triaxial accelerometers. Its recognition accuracy has no significant gain for a sampling frequency above 20Hz. The sensor used for this project, SHIMMER3 [1], has a sampling frequency of 50Hz, although it is downsampled to 16Hz afterwards.

The optimal place to carry the sensor is inside the trousers pocket[18]. Nevertheless, it depends on the activities to be recognized. The position selected for the patients of this study is their lumbar area. This position was decided because most of the basic activities can be recognized accurately and patients find it comfortable.

B.2.3 Activity classification: HMM

A Hidden Markov Model (or HMM) is a highly capable statistical tool for modeling time series. In a HMM we consider that the observations are independent given the hidden process that generates the sequences[19], and any given state only depends on the previous state.

In this work, each monitored activity corresponds to a different, independent, HMM[22]. The number of states of each HMM becomes a design parameter, in this work we employ five states per activity, following the model described in [20]. For every data sequence, we obtain the probabilities for each activity given the model using the Forward-Backward algorithm [21].

B.3 Objectives of the project

The objective of the thesis is to provide the means for a successful analysis of human behavior based on experimental data samples recollected from psychiatric patients. Doctors are interested in keeping track of their daily activity patterns.

The data sampled in each time instant, is interpreted and transformed into five probabilities that correspond to the five activities considered in this project (section 4.2). In other words, for each time instant there are five activity probabilities. As these samples are the result of prior sensor data processing [22][20] they are assumed to be valid and will not be modified in any way.

A graphical user interface is proposed to ensure such performance and flexibility, allowing to introduce the data selection parameters and any of the three main representations offered: day intensity, day evolution and hour details.

B.4 Description of the daily activities database

A total of 25 patients of the psychiatric wing of Fundación Jiménez Díaz have been assigned a random number with four to six digits in order to keep anonymity.

Within each patient's folder, it is provided a subfolder corresponding to each day recorded and the subfolder name establishes the date and starting time of each session. In each day subfolder, patient activity data is contained in a single ".mat" file, which contains a data structure that contains the probabilities of a patient performing one of the five activities considered.

For this project, the registered activities are five:

- **1st row,** running. Energetic activity, running or walking fast.
- **2nd row,** walking. Device is upright and detecting rhythmic movements consistent with walking.
- **3rd row,** standing. Patient is at upright position, small movements give away the activity (shifting weight from one leg to another).
- **4th row,** sitting. Patient is at upright position and there is no movement.
- **5th row,** lying down. Device is at horizontal position, so patient must be lying mostly static and not moving much.

B.5 Data processing

B.5.1 Data storage

Initially, the application will need to find the original patient data. It will ask the user to select the folder in which the data for all patients is contained.

The user will then introduce the *patient_ID*. This *patient_ID* is concatenated to the previous data path, obtaining as a result the full path corresponding to the specified patient. The user must also specify the date of interest. With these parameters the program will select the proper daily activity pattern data file and it will be subsequently loaded into a matlab cell unit.

Due to the patients' schedule, the proposed time frame is from eight o'clock until midnight with a time step of fifteen minutes. The input recording of the patient will be ordered and stored in this way.

We have decided to employ cell units to store patient data instead of matrices due to its simplicity and ordered manner. Cell units allow the storage of multiple data types of different dimensions simultaneously. For this project, we are interested in storing matrices as the elements of the cell. If for a determined time instant there is no input data, the cell unit elements equivalent to that time instant will be null, which means there will be no additional matrices full of zeros consuming resources.

Let us take into account the volume of data considered. If we consider a daily activity pattern file starting at eight o'clock and lasting until midnight, we obtain a total time range of sixteen hours. As the sampling frequency is 16Hz the sampling period is 0.00625s.

A total of 921600 samples have been gathered and for each sample taken there is a probability associated to each possible state; that is a total 4608000 probabilities for a single day for a single patient.

A day data file is around 24MB in size. For this project we have available around 300 days for a total worth of near 6GB of data.

B.5.2 Data classification and selection

First, the folder corresponding to *input_date* has its daily activity pattern file loaded into a cell. To achieve the correct order of the samples, we need to be very careful with the starting time of the input sample. An auxiliary function acts as translator between an hour and its column position.

The resulting cell is returned.

Then, another function returns the matrix containing the sampled statistic means of every time step of the input patient data. Within this function, daily activity pattern files are simplified enough to obtain a representation in matlab.

The last of these functions returns the vector containing the mode of every minute of the input patient data, also simplifying activity pattern files for visualization in matlab.

B.6 Graphical user interface

Let us consider the following example:

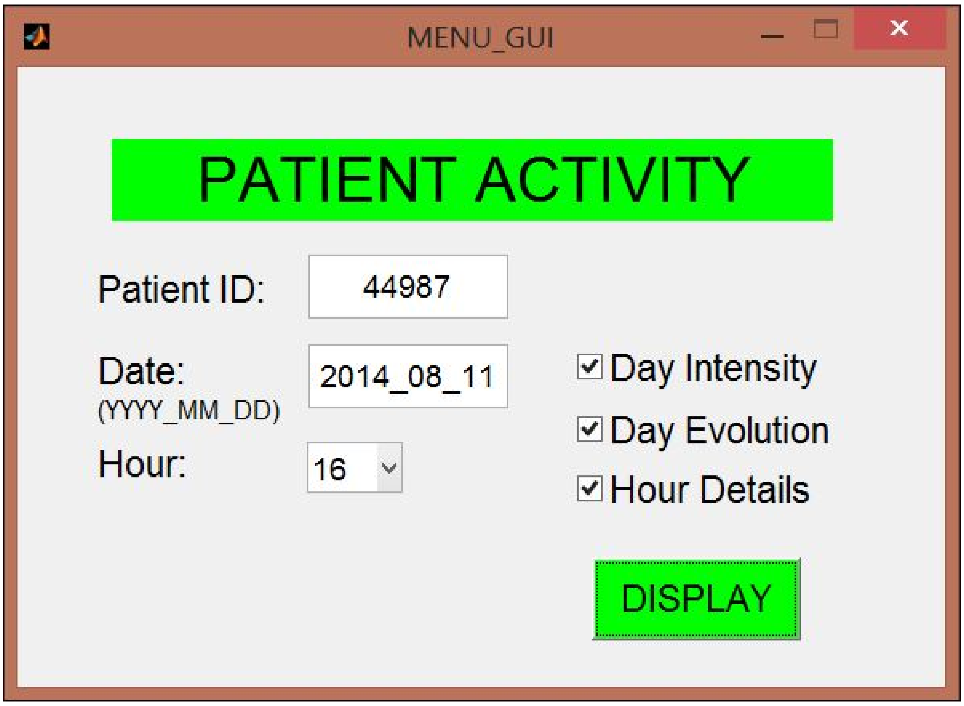


Figure 26: Example of the input data in the GUI.

There are three available representations of the daily activity pattern file:

- **Day Intensity**, as seen in Figure 27, represents the intensity of the activities performed by the patient. It's a temporal plot of the sampled probability means of the activities. As the hue gets more intense, the mean for the activity increases. The x-axis represents the time of the day from eight o'clock until midnight in steps of fifteen minutes (only sharp hours are labeled for reference). The y-axis represents the activities performed by the patient.

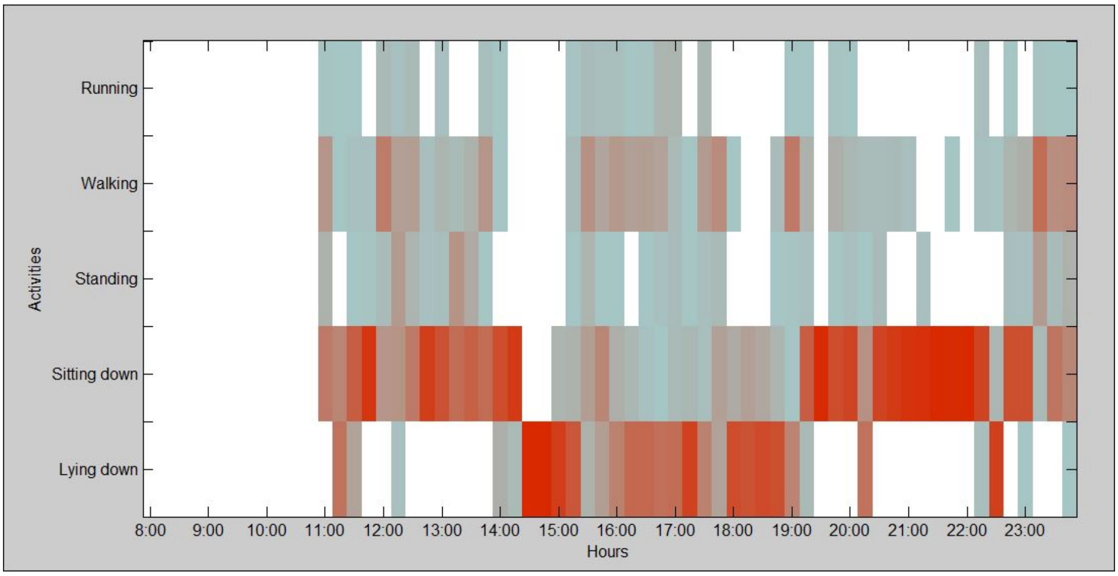


Figure 27: Day intensity graph for the example proposed.

- **Day Evolution**, as seen in Figure 28, represents the evolution of the activities performed by the patient along the day. It's an area plot of the sampled probability means of the activities. A darker hue stands for the more energetic activities, being "running" (high activity) the maximum with the darkest color. In the same way, a lighter hue suggests a minimum activity.

  The x-axis represents the time of the day from eight o'clock until midnight in steps of fifteen minutes (labels are only shown for full hours). The y-axis represents the accumulated probability of the activities performed by the patient, with high activity at the bottom and adding lower and lower on top of it. They will add up to one (100%), as should be expected.

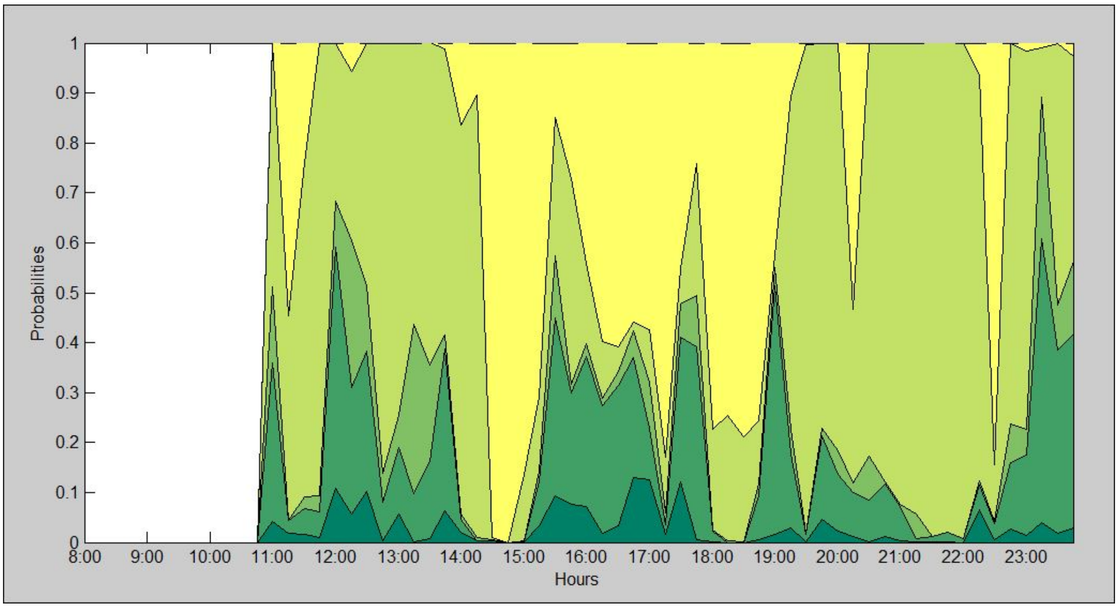


Figure 28: Day evolution graph for the example proposed.

- **Hour Details**, as seen in Figure 29, represents the activities performed by the patient. It's an area plot of the modes of the activities. The x-axis represents the minutes of the selected hour, from one until sixty in steps of one minute. The y-axis represents the activities performed by the patient.

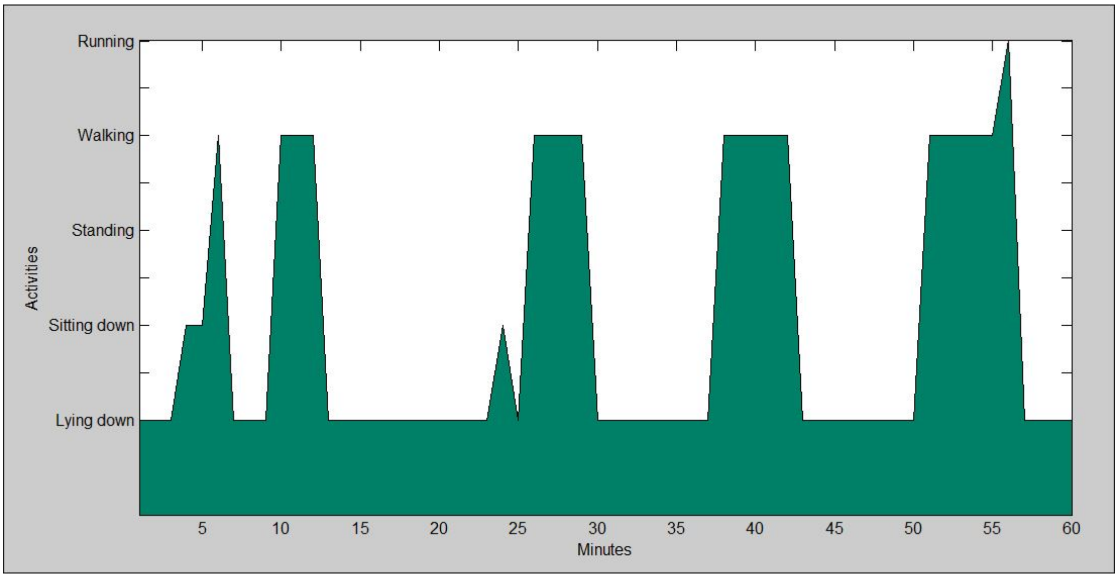


Figure 29: Hour detailed graph for the example proposed.

B.7 Alternatives

Let us discuss some possible alternatives to the work done:

- Data processing could be performed in other environment besides matlab. Number-crunching programs such as R or even Excel are an option to consider. Also, mathematical libraries in compiled languages like C could be used.
- Once data has been processed for presentation, graphs can be generated with specialized libraries for more appealing displays for the end user (like plot.ly within matlab). Other graphical libraries and compiled languages could be used. Even Excel could be also considered for display.
- By using a website, the end user interface would be easier to create and modify. All data processing and graphic images would be processed by the web server providing non-local access to the application. This processing in the background would allow independent programs for the different tasks, like file handling, data processing and visualization. Of course, this rises the concern for patient data protection. Access control system must be implemented.

B.8 Future work

The next step of the project is to develop the final application that will be used by medical personnel to implement all work done in this project. This final application should be accessible from either computer, tablet or smartphone.

However, a fundamental step is to confirm whether the proposed graphical representations are useful to doctors, with the aim of establishing a level of adequacy of these representations of patient behavior.

Patient feedback would also be a great option to consider. Via messages on the mobile app on their smartwatches or smartphones, patients could receive the aforementioned reports with their progress or statistics of the day. This feedback could be modulated by some parameters introduced by the doctors, in order to motivate them (patients) or limit them (athletes).

An example of these messages could be "You have walked a total of __ steps! Well done, you only need __ to reach your goal!".

References

[1] Shimmer. About us. [Online]. Available: http://www.shimmersensing.com/about-us/

[2] Y. Jia, "Diatetic and exercise therapy against diabetes mellitus," *Second International Conference on Intelligent Networks and Intelligent Systems*, pp. 693–696, 2009.

[3] J. Yin, Q. Yang, and J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 8, pp. 1082–1090, 2008.

[4] D. Lawrence, C. Holman, and A. Jablensky, "Death rate from ischaemic heart disease in Western Australian psychiatric patients," *British Journal of Psychiatry*, vol. 182, pp. 31–36, 2003.

[5] M. Berren, K. Hill, and E. Merikle, "Serious mental illness and mortality rates." *Hospital and Community Psychiatry*, vol. 45, pp. 604–605, 1994.

[6] E. Martinsen, "Benefits of exercise for the treatment of depression." *Sports Medicine*, vol. 9, no. 6, pp. 380–389, 1990.

[7] M. Farmer, B. Locke, E. Mościcki, A. Dannenberg, D. Larson, and L. Radloff, "Physical activity and depressive symptoms: the NHANES I Epidemiologic Follow-up Study." *Am J Epidemiol*, 1998.

[8] C. R. Richardson, G. Faulkner, J. McDevitt, G. S. Skrinar, D. S. Hutchinson, and J. D. Piette, "Integrating physical activity into mental health services for persons with serious mental illness." *PSYCHIATRIC SERVICES*, vol. 56, no. 3, pp. 324–331, 2005.

[9] M. Mellion, "Exercise therapy for anxiety and depression. 1. Does the evidence justify its recommendation?" *Postgrad Med*, 1985.

[10] S. A. Paluska and T. L. Schwenk, "Physical activity and mental health," *Sports Medicine*, vol. 29, no. 3, pp. 167–180, 2000.

[11] S. Petruzzello, D. Landers, B. Hatfield, K. Kubitz, and W. Salazar, "A meta-analysis on the anxiety-reducing effects of acute and chronic exercise. Outcomes and mechanisms." *Sports Medicine*, 1991.

[12] G. Faulkner and S. Biddle, "Exercise as an adjunct treatment for schizophrenia: a review of the literature." *Journal of Mental Health*, vol. 8, pp. 441–457, 1999.

[13] S. Weyerer and B. Kupfer, "Physical exercise and psychological health." *Sports Medicine*, 1994.

[14] Óscar D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE COMMUNICATIONS SURVERYS AND TUTORIALS*, vol. 15, no. 3, pp. 1192–1209, 2013.

[15] J. Iglesias, J. Cano, A. M. Bernardos, and J. R. Casar, "A ubiquitous activity-monitor to prevent sedentariness," *IEEE Conference on Pervarsive Computing and Communications*, 2011.

[16] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," *Proc. International Workshop on Wearable and Implantable Body Sensor Networks*, 2006.

[17] T.-P. Kao, C.-W. Lin, and J.-S. Wang, "Development of a portable activity detector for daily activity recognition," *IEEE International Symposium on Industrial Electronics*, pp. 115–120, 2009.

[18] Z.-Y. He and L.-W. Jin, "Activity recognition from acceleration data using ar model representation and svm," *International Conference on Machine Learning and Cybernetics*, vol. 4, pp. 2245–2250, 2008.

[19] P. Blunsom, "Hidden markov models." *Lecture notes*, 2004.

[20] B. Florentino, N. O'Mahony, and A. Artés-Rodríguez, "Hierarchical dynamic model for human daily activity recognition," *Proceedings of BIOSIGNALS*, 2012.

[21] L. R. Rabiner. A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. [Online]. Available: http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/tutorial%20on%20hmm%20and%20applications.pdf

[22] V. Elvira, A. Nazábal-Rentería, and A. Artés-Rodríguez, "A novel feature extraction technique for human activity recognition," *IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 177–180, 2014.

[23] C. Han, S. Kang, and N. Kim, "Implementation of HMM-based human activity recognition using single triaxial accelerometers," *IEICE Transactions*, vol. 93, pp. 1379–1383, 2010.

[24] T. L. project. LaTeX the product. [Online]. Available: http://www.latex-project.org/

[25] COIT. Verificación de proyectos de ICT. [Online]. Available: http://www.coit.es/web/ejercicio_profesional/svc/dv/2016_verificacion_ict.pdf

[26] Mathworks. Matlab: The language of technical computing. [Online]. Available: http://www.mathworks.com/products/matlab/index.html?s_tid=gn_loc_drop